

1. A computer-implemented method to search for data responsive to first and second query concepts, comprising:

receiving a first set of expanded results generated from a first query concept by utilizing one or more data sources;

5 receiving a second set of expanded results generated from a second query concept by utilizing the one or more data sources; and

determining an intersection set of documents from the first and second sets of expanded results, such that a relationship can be determined between the first and second query concepts from the intersecting set of documents.

10

2. The method of claim 1 wherein the relationship between the first and second query concepts is explained by determining for each document those concepts related to the document from a larger concept set, the larger concept set including expansions of the first query concept and the second query concept.

15

3. The method of claim 1 wherein a first relevance score is assigned to the first set of expanded results and a second relevance score is assigned to the second set of expanded results and wherein a composite relevance score is assigned to the intersection set of documents.

20

4. The method of claim 3 wherein the composite score is assigned by multiplying the first and second relevance scores.

5. The method of claim 1 wherein the documents are filtered by a relevance score.

6. The method of claim 1 wherein the expanded results are generated by:

5 defining a first set of documents relevant to the query concept, the first set of
documents being a subset of a collection set of documents;
 building a first histogram of features from the first set of documents; and
 selecting features for an expanded feature set by comparing the first histogram of
features with a second histogram of features from the collection set of documents.

10 7. The method of claim 6 wherein the features in the second histogram are a baseline
expansion feature set and the features for the expanded feature set are selected by
removing features from the baseline expansion feature set based on how often the
features appear in the first histogram.

15 8. The method of claim 7 wherein the baseline expansion feature set is generated by
training on a random data sample.

9. The method of claim 6 wherein the expanded feature set is ranked by expected
entropy loss.

20

10. The method of claim 6 wherein concept constraints are applied to the expanded
feature set.

11. The method of claim 6 wherein a feedback scoring function is applied to results generated from the expanded feature set.

12. The method of claim 11 wherein the feedback scoring function assigns a fixed score to each feature and where feature can be assigned different fixed scores.

13. A computer-implemented method for automatic query expansion comprising:
defining a first set of documents relevant to a first query concept, the first set of documents being a subset of a collection set of documents;
building a first histogram of features from the first set of documents; and
selecting features for an expanded feature set by comparing the first histogram of features with a second histogram of features from the collection set of documents.

14. The method of claim 13 wherein the features in the second histogram are a baseline expansion feature set and the features for the expanded feature set are selected by removing features from the baseline expansion feature set based on how often the features appear in the first histogram.

15. The method of claim 14 wherein the baseline expansion feature set is generated by training on a random data sample.

16. The method of claim 13 wherein the expanded feature set is ranked by expected entropy loss.

5 17. The method of claim 13 wherein concept constraints are applied to the expanded feature set.

18. The method of claim 13 wherein a feedback scoring function is applied to results generated from the expanded feature set.

10 19. The method of claim 18 wherein the feedback scoring function assigns a fixed score to each feature and where feature can be assigned different fixed scores.

20. A computer-readable medium storing instructions to search for data responsive to first and second query concepts, the medium comprising instructions for:

15 receiving a first set of expanded results generated from a first query concept by utilizing one or more data sources;

receiving a second set of expanded results generated from a second query concept by utilizing the one or more data sources; and

20 determining an intersection set of documents from the first and second sets of expanded results, such that a relationship can be determined between the first and second query concepts from the intersecting set of documents.

21. The computer-readable medium of claim 20 wherein the relationship between the first and second query concepts is explained by determining for each document those concepts related to the document from a larger concept set, the larger concept set including expansions of the first query concept and the second query concept.

5

22. The computer-readable medium of claim 20 wherein a first relevance score is assigned to the first set of expanded results and a second relevance score is assigned to the second set of expanded results and wherein a composite relevance score is assigned to the intersection set of documents.

10

23. The computer-readable medium of claim 20 wherein the expanded results are generated by:

defining a first set of documents relevant to the query concept, the first set of documents being a subset of a collection set of documents;

15

building a first histogram of features from the first set of documents; and

selecting features for an expanded feature set by comparing the first histogram of features with a second histogram of features from the collection set of documents.

20

24. A computer-readable medium storing instructions for automatic query expansion, the medium comprising instructions for:

defining a first set of documents relevant to a first query concept, the first set of documents being a subset of a collection set of documents;

building a first histogram of features from the first set of documents; and

selecting features for an expanded feature set by comparing the first histogram of features with a second histogram of features from the collection set of documents.

5 25. The computer-readable medium of claim 24 wherein the features in the second histogram are a baseline expansion feature set and the features for the expanded feature set are selected by removing features from the baseline expansion feature set based on how often the features appear in the first histogram.

10 26. The computer-readable medium of claim 25 wherein the baseline expansion feature set is generated by training on a random data sample.

27. The computer-readable medium of claim 24 wherein the expanded feature set is ranked by expected entropy loss.

15 28. The computer-readable medium of claim 24 wherein concept constraints are applied to the expanded feature set.

29. The computer-readable medium of claim 29 wherein a feedback scoring function is applied to results generated from the expanded feature set.

20 30. The computer-readable medium of claim 29 wherein the feedback scoring function assigns a fixed score to each feature and where feature can be assigned different fixed scores.